International Review of Social Sciences and Humanities
Vol. 7, No. 1 (2014), pp. 219-231
www.irssh.com
ISSN 2248-9010 (Online), ISSN 2250-0715 (Print)

An Exploration into the Evolution of Working Memory Models

Christopher A. Was
Lifespan Development and Educational Sciences, Kent State University
405 White Hall, Kent State University, Kent, OH, USA-44242
E-mail: cwas@kent.edu

(Received: 21-11-13 / Accepted: 28-12-13)

Abstract
Models of working memory (WM) have undergone an interesting, although slow evolution. From the early concept of primary memory to the view of long-term working memory the concept has undergone some important changes. There is still a great deal of dissent concerning of what exactly WM is composed and/or the processes involved. In this review of the literature concerning WM, an area of research has been uncovered that could lead to significant advancement of the understanding of WM processes as well as information processing and complex cognitive process, such as comprehension.

Keywords: Working Memory, Models of Working Memory.

1. Introduction to Working Memory Literature

The research leading to the construction of working memory (WM) models has taken many forms, using various techniques. Still, there are many unanswered questions. The most basic of which is, “What is working memory?” Through the current review of different models and conceptualizations of WM a number of more specified, and perhaps more intriguing, questions worth pursuing as a basis for research have arisen. The goal of this manuscript is to enlighten the reader in matters of WM and anchor questions in the literature regarding WM.

When first investigating the different models of WM one might seek to simply compare and contrast a few theoretical models and have a relatively complete understanding of what WM entails. One may even naively believed that after a thorough literature review a good understanding of the meaning of the term *working memory* could be derived. The reader might hope to understand if WM is a capacity or a set of processes, whether it is solitary or based on multiple components and whether it is domain specific or a general feature of memory.

What one finds is that all of these questions have been addressed in the research literature and that there is evidence for both sides of each of the above dichotomies. As it turns out, it is more efficient to describe the evolution of WM theory and then to analyze the research that has perpetuated that evolution.

2. Primary Memory

In his writings, the British philosopher John Locke referred to *retention* as a faculty of the mind. According to Locke [1], retention occurs in two ways.

“First by keeping the idea which is brought to me into it, for some time into view, which is called contemplation. The other form of retention is the power to retrieve again in our mind those ideas which after imprinting have disappeared, or have been as it were laid aside out of sight…This is memory which as it were the storehouse of ideas [1].”

The similarity between Locke’s two forms of retention and the concepts of primary versus secondary or short-term versus long-term memory will become quite clear in this discussion. These ideas are still the basis of most theorists’ views of the simple distinctions between memory systems.

William James [2] coined the term *primary memory* to describe the “specious present.” The term *specious present* is often interpreted as our awareness of thought or consciousness. James used the word *specious* to describe primary memory because he felt that one could not be certain about the accuracy of one’s thoughts. Our personal interpretations have an effect upon the precision of the information that is held in consciousness. In 1965, Waugh and Norman [3] revived the term primary memory as well as the term secondary memory. The term primary memory described a theoretical system responsible for short-term storage that was limited in capacity, and in which information is easily displaced by new information. The original information is displaced unless consciously held in primary memory through means of rehearsal. They also stated that rehearsal was the mechanism that transferred information from primary memory to secondary memory. Secondary memory refers to a hypothetical long-term memory system. The tasks used to measure short-term memory likely reflected processes involved in both primary and secondary memory, while delayed recall tasks most likely reflect only secondary memory. This view of the two basic memory systems is still an important construct in cognitive psychology. This is evident when one reads any introductory psychology or education textbook.

3. Short-Term Memory

The concepts of primary and secondary memory, though still important in terms of views concerning the limitations, capacities, and processes of the memory systems, eventually gave way to the concepts of *short-term* and *long-term* memory. It appears as though this change of terms came about because the connotations of the terms primary and secondary are somewhat misleading. Waugh and Norman [3] believed that information or memories needed to pass through the primary memory before they could be stored in secondary memory. However, this view does not account for memories that are stored in secondary memory that were not rehearsed in primary memory. Examples of these types of memory are episodic memory, procedural memory and implicit memory. Therefore, there was a shift from use of the term primary memory to the use of short-term memory and the name change indicated how the view of this memory system shifted. No longer seen as the “first” memory system, short-term memory came to be seen as a place where information was briefly held. Secondary memory now came to be seen as long-term memory where information is stored for a long period or even indefinitely.

4. Working Memory

Atkinson and Shiffrin [4] described the short-term store as a combination of storage and control processes. The term short-term memory (STM) referred to experimental situations designed to measure short-term store. Their research led to the theory regarding a control process beyond verbal rehearsal accounting for other memory strategies. In order to better describe STM and the control processes involved Atkinson and Shiffrin introduced the term *working memory*.

The Atkinson and Shiffrin [4] model of working memory (WM) describes the storage and processing of auditory, verbal and linguistic information. Atkinson and Shiffrin proposed that WM contained an auditory, visual and language (A-V-L) storage buffer and that storage and control processes utilize a shared resource. Put differently, there is a limited amount of resources available to WM, if the resources are being used to process information, then there will be less to devote to the storage of A-V-L information. This led to the view that WM is a flexible yet limited system.

Atkinson and Shiffrin also raised the possibility of a visual buffer however, there was not a great deal of evidence for this aspect of WM and the notion of multiple temporary buffers was set aside. The model came to be viewed as housing a single system for processing and storage of information.

This new view of the processing of information occurring in WM intrigued many researchers. Craik and Lockhart [5] elaborated the idea of WM as a processor. Their theory stated that the level at which information was initially processed determines ease and accuracy of later recall. This theory came to be known as the Levels of Processing Approach. In the Levels of Processing Approach WM is the cognitive processor. The processes in this system include rehearsal, but also include semantic and lexical judgments, as well as phonological and graphical decisions. Craik and Lockhart felt that the deeper levels of processing (e.g., when semantic and lexical judgments are involved) led to better retrieval. Deeper levels of processing create more connections to the information. For example, if the information being processed in WM is an elaboration of previous knowledge the new information may be better organized once encoded in long-term memory.

It may not be widely recognized that the levels of processing approach made a clear distinction between WM and long-term memory (LTM), but it most certainly did [6]. WM is the location where the processing of the information takes place. LTM is the location where information is stored. The relevance of the computer analogy and the connection to the information processing models is quite evident in this statement. In the computer analogy of the information processing model WM is seen as the central processing unit (CPU) where the data is input, manipulated and then passed to the hard drive (LTM) for permanent storage and later retrieval. This analogy also connects us to the concept of *activated LTM* in which WM is seen as the currently activated portion of LTM. This will be discussed in detail later in this manuscript.

4.1 Measuring Working Memory: The Reading Span

Daneman and Carpenter [7] designed a task that they reported to be a measure of WM span. This task purported to tap into both the processing and storage capacities of WM. The task involves the subject reading a number of unrelated sentences and then being asked to recall the last word in each sentence. The number of final words the subject is able to recall successfully is said to be the subjects “reading span.” This reading span according to Daneman and Carpenter reflects a limited capacity of the ability to process text. Daneman and Carpenter assumed the system that is constrained by these limitations is that of WM.

Just and Carpenter [8] expanded this model of WM in a series of studies in which they correlated reading span with reading times. King and Just [9] preceded the Just and Carpenter study by measuring reading times for simple sentences and sentences that are more complex. Based on this design for measuring reading times, Just and Carpenter found that subjects with low reading spans took longer to read complex sentences than subjects with high reading spans. Also high reading span subjects times were less affected by concurrent memory load tasks. The implication is that high reading span subjects have larger WM capacity than shorter span subjects and have a greater amount of resources available for processing and storage.

This led Just and Carpenter [8] to describe a theory of cognitive capacity limited by an available amount of activation. This budget of activation can be allocated flexibly but the amount of budgeted activation that has been allocated must, for any new processing or storage to occur, be reduced in some existing activation. For example, during reading tasks for new representations of later text to occur earlier text representations must be eliminated to allow for new representations of the later text. The reading span task is thought to measure the activation budgeted in reference to language comprehension.

The Just and Carpenter [8] theory focuses on the idea that reading span is a measure of WM capacity specifically in regards to language processing. However, several other researchers have shown that reading span correlates highly with other tasks. This is especially true of tasks that require processing and storage of information.

Baddeley, Logie, Nimmo-Smith and Brereton [10] devised a task, using sentences that were much less complex than the Daneman and Carpenter [7] sentences, to create a variation of the reading span task. The sentences in the task either made sense (i.e., could be judged as true) or were non-sense sentences (i.e., could be judged as false). Participants were required to judge whether the sentences made sense or not as each sentence was presented. After the sequence of sentences was complete, participants were required to recall either the grammatical subjects or objects of the sentences. Participants were not told in advance whether the subjects or objects would be the items of recall. Because of this a heavy memory load was placed on the participants as well as a processing requirement based on the judgment of sentence content.

Baddeley, et al. [10] found that this form of the reading span task was highly correlated with measures of comprehension. Case, Kurland and Goldberg [11] had devised a “counting span” task and Baddeley, et al. found that this task also highly correlated to the revised reading span task. Logie [7] claims that this is evidence that reading span, as a WM measure, may not only be applied to language processing, but is more likely a measure of a general purpose system for processing and temporary storage. Recent investigations have led to the construction of a number of *complex span* tasks designed to capture the storage and processing of information in WM [12. 13].

5. The Multiple Components Model

Baddeley [14] and other researchers have argued that this general-purpose system is comprised of multiple components. The most often cited research is that of Baddeley and his colleagues Hitch and Logie [15, 16]. Baddeley refers to his model as a multiple component model of WM. According to this model, WM consists of multiple specialized components that allow humans to comprehend, interact with the environment, represent information, and retain information about the immediate past and act on certain goals.

The three major components of the original model are the Central Executive, the Phonological Loop, and the Visual-spatial Sketchpad. The two temporary memory systems or slave systems, the phonological loop and the visual-spatial sketchpad, are used to maintain memory traces that overlap with sensory memory using rehearsal in the phonological loop and image generation in the visual-spatial sketchpad. The central executive controls features of WM such as the coordination of the two slave-systems, attention, focus, and attention switching, as well as activation of long-term memory [17]. Importantly, Baddeley does not believe that the central executive is used for any temporary storage of information.

Baddeley contends that each of the slave-systems can also be broken down into two subsystems. The phonological loop, for example can be fractionated into a passive phonological store and an active rehearsal process. The phonological store represents material in a phonological code that decays with time unless the material is processed through rehearsal. Baddeley sites not only neuropsychological findings [14], but also nueromimaging findings [18] to support his claim. The visual-spatial sketchpad is also fractionated into passive and active subsystems. Logie [6] refers to these as the passive visual cache and an active inner-scribe.

In more recent versions of the multiple components model [16, 19] Baddeley includes an episodic buffer. In these versions of the model, the episodic buffer is a storage system that uses multimodal elements to coordinate the slave systems with LTM. Baddeley [19] stated that when considering recall of prose the episodic buffer can be compared to Ericsson and Kintch's [20] proposed LTWM. This leaves the central executive as an attention system not limited to WM processes.

The usefulness of the visual-spatial sketchpad and the phonological loop is that the passive aspects overlap with sensory memory and perception to set up a more durable memory trace (although still very short in duration). These memory traces can then be acted upon, or not, by the active rehearsal mechanisms. It appears that rehearsal is not a necessary condition for WM via the central executive to process information. However, rehearsal can be used to prolong access to information as well as enhance performance.

At this point the reader might ask, "How, if the capacity of the two subsystems is very limited and the central executive has no storage capacity, can one hold large amounts of information?" More importantly, "How can such a small capacity account for the ability to process great amounts of information in highly skilled tasks or the processing of experts in a specific domain?"

Baddeley [17] stated that WM can utilize temporary storage in systems other than the two slave-systems. He discussed the recency affect in LTM as a relatively automatic process coupled with specific and active retrieval processes. That is, information that was most recently stored in LTM is the first to be recalled. The recency effects, as well as the primacy effect, have been demonstrated in LTM studies as well as WM. In WM the majority of evidence is based on studies involving word lists. This is also the case for evidence regarding the primacy effect in LTM, but Baddeley contends that this effect can be extended beyond just simple lexical items to high-level semantic structures including complex schema and procedural memory. The inclusion of the episodic buffer also helps to explain the expansion of capacity for complex cognitive processes.

Ericsson and Kintsch [20] claimed that the mechanisms of short-term memory (storage in WM) can account for WM in unfamiliar activities, but cannot account for the necessary storage capacity for WM in skilled complex activities. According to Ericsson and Kintsch, storage in WM can be increased and is one of many skills individuals attain during skill acquisition. This increase in WM is aided by activation of procedures and schema stored in long-term memory. Ericsson and Kintsch refer to this as long-term working memory (LTWM). LTWM, also known as expert memory, is domain specific and limited to the skilled activity in question. The fact that LTWM is acquired implies that there are significant individual differences and differences in the skills for which it can be acquired.

Although LTWM is domain specific, in those tasks in which it has been acquired the expert will have greater storage resources for the information to be processed because the slave-systems will be free from the storage of the procedures necessary to perform the task. The research involved in developing these hypotheses led Baddeley and others to the conclusion that the central executive does not posses a storage capacity, but that WM as a whole employs the slave-systems and other resources for storage. Duff and Logie [21] designed a study to determine if this hypothesis could be backed up empirically. The most important finding of this study was that a demanding processing task had virtually no effect on the capacity of storage.

The major problem of hypothesizing the central executive as only a control structure is that many will regard the central executive as a homunculus. Baddeley [19] himself has acknowledged this shortcoming. Baddeley's early work focused on the central executive's function of controlling the two slave-systems. In attempts to clarify the central executive's role, his later work began to focus on the central executive's other functions including the capacity to focus attention, to switch attention, and to activate long-term memory.

Baddeley and Logie [16] acknowledge that WM is closely related to attention and that the central executive is often seen as an attentional system. Baddeley even states that a better term might be *working attention*. Baddeley and Logie stated that is important to note that attention is a non-unitary concept.

6. Attention and Working Memory

Posner [22] distinguishes between three different attentional systems that control orientation, alertness, and high-level attention. Posner's evidence is based on the study of anatomical structures. Without going into a discussion of the anatomical structures involved, the three systems involved include: (1) the system in control of orientation which is based on a reflex system or automatic process, (2) the system in control of alertness refers to a system used for focus to control elicited responses, and (3) the systems in charge of high-level attention are truly concerned with what an individual is focusing on in WM. Even if we consider these processes to be under the control of the central executive, we have not gone beyond the idea that the central executive is a mere homunculus making decisions about to what to we should pay attention. Posner provided detail about the affects environmental stimuli and spread of activation patterns will have on the three attentional systems. Baddeley and Logie [16] neglected to fully incorporate these systems into the central executive. Baddeley [19] argued that attention processes in WM allow the organism to choose from available options and to select a particular strategy. The question remains, how does this differ from the Homunculus?

7. The Embedded Processes Model

There is a great deal of evidence to support the existence of a phonological loop and a visual-spatial sketchpad. However, how does one focus attention or decide which information to retrieve from long-term memory? Cowan [23, 24] attempts to answer these questions concerning attention and activation in his Embedded Processes Model of WM. The most significant features of the embedded processes model are that information in WM comes from "hierarchically arranged faculties comprising long-term memory, the subset of long-term memory that is currently activated and the subset of activated memory that is currently the focus of attention. [23, p. 62]" According to Cowan the focus of attention is controlled by a voluntary process located in the central executive and involuntary processes that are controlled by an attention orienting system.

On first review of Cowan's embedded processes model there is a great lack of information concerning the processing of new information. However, as one reads further explanations of how new information is handled by WM become clear. It is this author's opinion that this continues to be a weakness of his model, however I feel that his views on LTM activation are important. Therefore, I will focus on this aspect of his model.

Cowan [23] defines WM as any processing mechanism used for temporary storage, availability, and processing of information. Hence, Cowan's views of WM are more concerned with the processes of WM and not the actual mechanisms. This is an important distinction in Cowan's model and it is helpful to consider Cowan's views concerning the activation of information from LTM to understand this distinction. Cowan's views of activation are important in many models of WM that consider the contents of WM the activated portions of LTM. According to Cowan, [24] different WM processes have different capacities. The focus of attention, for example, has a capacity limitation. That is, only a limited amount of information can be held in the focus of attention, whereas activation of information is time limited. If information that has been recently activated is not attended to the activation fades. Therefore, if information relevant and irrelevant to a specific task is activated from LTM and the irrelevant information is attended to, there may only be enough "room" in the focus of attention for that irrelevant information. By the time this incorrect information has been processed the relevant information may not longer be activated.

Attention in this model is seen as the processing of some activated information to the exclusion of other information. It is not possible to attend to information without it being active, but it is possible to activate information without it being attended to. This raises pertinent theoretical and experimental questions: 1) How likely is it that irrelevant information that has been activated will come into the focus of attention? 2) If the contents of WM are the activated portion LTM, how does the control of attention differentiate between relevant and irrelevant information? 3) If focus of attention is a voluntary process, are there individual differences in the ability to control attention? This last question has undergone a good deal of research, but is still considering.

In an attempt to answer, similar questions Conway and Engle [25] propose the Resource-dependent Inhibition model of WM. According to Conway and Engle most psychologists have accepted that memory consists of two main compartments. However, they state that new conceptions of memory assume that WM is simply activated long-term memory. They also state however, that the relationships among short-term, working and long-term memory are probably much more complex. There is an important distinction lacking in this viewpoint. That is the distinction between WM and short-term memory. I would agree with Conway and Engle that the line between long-term memory and another memory system as the activation of long-term has become blurred. However, this is the only article in which the memory system considered the active portion of LTM is short-term memory. That distinction is usually reserved for WM. In attempt to clarify for the sake of this particular article Conway and Engle, refer to the active portion of memory as primary memory, and the inactive portion of memory as secondary memory. These refer to James' terms for the two memory systems and reverting to these terms adds more to the confusion of these concepts than clarification.

8. Activation of Long-Term Memory

Conway and Engle [25] operationalize memory activation based on Anderson's ACT* model [26]. This model will be discussed in greater detail later in this review, however for purposes of discussing this article it is necessary for the reader to have a basic understanding of how ACT* describes activation of memory.

ACT* assumes that WM is LTM that has been activated beyond some critical threshold of excitation. Activation is considered a resource with limitations that occurs automatically and spreads among related activated concepts. As the activation of a concept reaches the critical threshold, the concept becomes accessible. As activation increases so does accessibility.

Conway and Engle [25] argued that primary memory is memory that has been adequately activated to be readily accessible and secondary memory to be memory that has not been activated to the critical threshold. Again, these operational definitions leave ambiguity regarding distinct memory systems. If, as the level of activation increases so does accessibility, how can one make a clear distinction between primary and secondary memory? According to this operational definition, would it not be the case that when a memory is activated beyond the critical threshold and then enters primary memory there are also levels or degrees of accessibility in primary memory? This would then render their attempts to distinguish retrieval from primary memory vs. secondary memory rather difficult unless they attempt to keep the level of activation in primary memory constant.

Conway and Engle [25] after review of the activation literature conclude that the general-capacity model is the best explanation of the relationship between WM capacity and LTM activation. They also conclude that this relationship is one in which WM is equivalent to the amount of activated LTM. The level of activation is that which will differ among individuals. This raises the question of whether the amount of activation varies within individuals. Cantor and Engle [27] contested that individual differences in the capacity of WM exist because of the available activation of LTM. If this statement is true, high and low capacity WM subjects should show different patterns of results in retrieval tasks. This would be the case if the specific amount and levels of activation could be controlled. The argument is not that a difference of retrieval from primary versus secondary memory cannot be found, but that a precise difference would be difficult to ascertain.

Conway and Engle [25] designed a series of experiments to determine the difference between primary and secondary memory retrieval. The second issues concerns whether WM capacity plays a role in memory retrieval. Once again in order to avoid detail Conway and Engle suspected that retrieval from primary memory consisted of a search of the learned material. Retrieval from secondary memory consists of a search for the list containing the specific item and then a search of primary memory of the activated list for the specific item. Once again, it appears that Conway and Engle are making a distinction between the process of WM and the storage capacity of STM.

The findings of this series of experiments are rather enlightening. First, the time to retrieve a set from secondary memory into primary memory did not vary with set size. This implies that activations of learned sets were in the form of set activation. Second, WM capacity had no effect on retrieval time from secondary memory. WM capacity did have an effect on the search time on the set of information that was active in primary memory but only when there was an overlap of set members was this effect found. Conway and Engle ask their readers why this might be the case. Their explanation is based on inhibition.

It is necessary at this point to discuss Conway and Engle's [25] findings in the context of their research questions. First, they asked if retrieval from primary and secondary memory is based on the same process. Citing Wickens, Moody and Dow [28], Conway and Engle expected retrieval from secondary memory to be a two-stage process, with the first stage being the retrieval and activation of information concerning the entire list of items containing the one to be retrieved. The second step would be the searching the list now in primary memory for the specific item. Wickens et al. (1981) claimed that the step necessary to retrieve the list from secondary memory would not vary in retrieval time whether the list consisted of two items or four items. Conway and Engle found that there was no difference in retrieval time if the list was 4, 6, 8, 10, or 12 items in length. However, there was a set size by delay of cue interaction if a set size of two was included in the analysis. It is unclear why this would occur and Conway and Engle did not address this question either. What they do conclude is that in set sizes of greater than four they affirm Wickens et al.'s conclusion that the process involved in retrieval from primary memory is different from that of secondary memory since set size has no effect on retrieval time from secondary memory, but does have an effect primary memory.

Another interesting conclusion of Conway and Engle [25] is that set size of four is below WM span and a set size of 12 is supra-span for the majority of their research participants. The fact retrieval time from secondary memory did not interact with group (that is high versus low span subjects) retrieval from secondary memory to making information available to primary memory must be an automatic process. They conclude this is evidence that activation of information in secondary memory is not a capacity limited process and that activation of information in secondary memory is in lists or chunks when applicable.

Conway and Engle [25] did find that WM capacity interacted with set size on retrieval time when membership in the sets of to be learned and retrieved information overlapped. High-span subjects were not as affected by overlap in the sets. One is led to conclude that interference is the cause of this interaction and that high-span subjects must be better at eliminating interference when searching for set information. What Conway and Engle conclude is that when sets overlap activation spreads to the irrelevant set. According to this view in the experiments in which no overlap occurred high and low-span subjects were relying on the automatic process of activating information in secondary memory. In the two experiments in which overlap was present activation spread not only to the relevant set but also to the irrelevant set of information. High-span subjects, who presumably have more available attention resources, used these resources to inhibit the activation of weaker irrelevant information. Low-span subjects who do not have the resources to inhibit the activation of irrelevant information allowed this information into primary memory and completed a serial search of both the relevant and irrelevant lists before being able to confirm the information they set out to retrieve.

Bjork [29] described inhibition as an important, adaptive memory mechanism. He argued that inhibition occurs in two distinct ways: (1) suppression, an active process that is directed at the to-be-inhibited information to enhance the possibility of achieving some specific goal and (2) blocking, occurs when an increase of activation in one bit of memorized information blocks or inhibits activation of other information. Bjork considered this an automatic process.

Conway and Engle [25] argued that suppression, because it is an active process must be resource dependent and that these resources would not be available if attention is otherwise being spent in other processes or if the individual has a smaller WM capacity to begin with. Blocking, on the other hand, is similar to activation in that it is an automatic process. Suppression must explain the difference between high and low-span subjects in the portion of their experiments where overlap between lists occurred. This view of inhibition as two separate processes, one automatic and the other a controlled limited capacity, are similar to Posner's [22] view of the separate mechanisms of attention. All of these views lead to some interesting research questions. First, what is the relationship between attention, activation, and inhibition? Second, are activation from primary or short-term memory and suppression truly controlled limited capacities?

Engle and his colleagues later expanded on these ideas to the point that they equated the importance of attention with general fluid intelligence [30]. Engle, et al. reflected on previous research findings and the questions that have influenced Engle's research. These questions are "what accounts for individual differences on measures of WM capacity?" and "What is measured by the complex tasks that is also important to higher-level cognitive tasks?"

To answer these questions Engle et al [30] described WM as a system that consists of the long-term memory traces that are currently activated above a threshold, the processes involved in achieving and maintaining activation, and controlled attention. If one recalls Conway and Engle's [25] view of WM as the portion of activated LTM we see that there has been a bit of evolution in Engle's view of WM. In his view, WM still consists of activated LTM, but there are now also some processes involved beyond just that activation. However, Engle has more recently defined WM capacity as *the capacity for controlled, sustained attention in the face of interference of distraction.* Individual differences in WM capacity then reflect differences in the ability to control attention during tasks that require controlled attention. Therefore, if WM consists of the portion of LTM that is activated above a threshold, then WM capacity is the ability to bring these activated memory traces into focus and hold focus on the task relevant memory traces in the face of distraction.

9. ACT-R

The ACT R theory of cognition is a fixed computational architecture that applies to all cognitive tasks [31]. In the ACT-R architecture one creates models for different cognitive tasks. The difference among the models is not the way in which information is processed, but instead it is the information initially stored in relation to the task. This initial knowledge includes the facts and skills that are necessary for the task to be performed.

In order to understand how the initial knowledge is stored we must examine the major aspects of ACT-R. The ACT-R system consists of both symbolic and sub-symbolic aspects. Knowledge is represented symbolically and processes that act upon knowledge are represented sub-symbolically. In order to discuss the ACT-R perspective of WM it is necessary to describe the two symbolic representations that make up knowledge in the ACT-R model.

Declarative knowledge is represented as a network of interconnecting nodes. This is a common representation of declarative knowledge in the memory literature. The ACT-R system representation of procedural knowledge is that of a set of productions. Each production represents a condition for action. The productions take the form of IF-THEN statements. IF a specific condition relevant to the task is present, THEN a specific action is to be taken.

The processing of declarative knowledge and procedural knowledge is directed by the current goal of the system (for example determining the product of the multiplication of two numbers). According to the ACT-R model the current goal contains the information in the focus of attention and is based on a declarative node structure. It appears that the contents of the focus of attention are established by previous processing, for example, upon completion of one production in the completion of a task the next production is set up. Environmental stimuli may also cause the switching of attention to different focus. This explanation of the focus of attention is reminiscent of Posner's [25] model of the attention mechanism or processes.

In order to explain how productions are selected for use Lovett and Anderson (1996) discuss the expected utility of a production. According to Lovett and Anderson, each production represents a separate unit of procedural knowledge and the environment regulates the processes acting on this knowledge. The process of selecting among several problem-solving activations is equivalent to selecting among several *production instantiations* (a production mapped onto the specifics of the current situation Lovett and Anderson, [32]). The selection of a production instantiation is based on the probability of success and the lowest expected cost to the system. The system determines cost by deciding which production actions will result in ending closest to the final goal as well as their probability of success. Therefore, ACT-R is more likely to select the production instantiation that is most likely to take the current state closest to the goal state.

ACT-R allows for the history of success to play a determining role. According to the ACT-R model a production is one unit of procedural knowledge and therefore the history of success information is stored at the production level. History of success is stored as the proportion of times that the production has led to success and failure therefore determining the probability of success.

Goals in the ACT-R architecture work as a filter for productions in that they are used to measure which productions have the highest probability of bringing the system closest to the final goal state. This filter uses the probability of success based on previous success as well as the utility of the production in goal attainment to determine which production should be initiated.

Once a production has been selected, retrieval related to this production is attempted. The production selected and the retrieval patterns with which it is associated affects retrieval, and the declarative knowledge connected to the current goal. Anderson and others have described the processes involved with the ACT-R architecture in terms of mathematical formulas[31] [33]. For this manuscript, I do not feel it is necessary to describe the formulas, but instead to discuss what they represent in terms of activation of memory and WM.

According to the ACT-R architecture each node in declarative knowledge has a base-level activation that influence its' accessibility. Each node is given its' base-level activation when it is first stored. Each time the node is activated after initial storage it receives a boost to its' base level activation. This boost can be thought of as learning. These boost decrease as a function of time, which can be considered forgetting. According to the model, nodes used more often are more likely accessed because they are more accessible. Reflecting on this leads us back to the Conway and Engle [25] study in which I questioned whether the amount of activation would affect the ability to retrieve information above the threshold. If the reader recalls, Conway and Engle operationalized their concept of activation for that study based on the ACT-R model, but did not allow for differing levels of activation.

Base level activation represents a node's accessibility in the ACT-R model. However, base level activation does not account for the effects of context on accessibility. According to the ACT-R model, one's goals represent the focus of attention and hence serve to create context effects, raising the activation of certain nodes of declarative memory relative to others. Past use of a fact, along with its relevance to the current goal determine accessibility of the fact. Therefore, familiar facts will be more accessible than unfamiliar facts and goal relevant facts will be more accessible that irrelevant facts. The equations involved have been used to successfully design and model experimental designs [34].

The processes involved in WM from the ACT-R perspective are the processes involved in the attentional activation system that differentially activates information relevant to the current goal context. More simply put, WM processes are those process that select which highly activated nodes of declarative knowledge on which to focus. It is apparent that both content and processing are important to the ACT-R perspective of WM and that the basic mechanism of WM is the spreading of activation which more strongly affects nodes of declarative knowledge that are closely related to the current goal.

Finally, because the discussion has included research questions relating to attention and inhibition I thought it pertinent to investigate how the ACT-R perspective views the relationship of WM and attention. In the literature regarding ACT-R and the earlier version, ACT*, little is provided regarding the role of attention. Lovett et. al., [31] stated that the notions of WM and attention are closely related in the ACT-R perspective. However, this relationship appears to be based in the hypothesis that both WM and attention are limited by the amount of source activation available. Specifically, the ACT-R architecture limits WM to the amount of attentional resources available. Several of the tasks used in research based on the ACT-R model have been dual-task situations in which attention is divided among two tasks and hence spread more thinly among a large number of goal nodes.

Lovett et al. [31] stated that attention often refers to an aspect of motivation or alertness suggesting that attention can consciously be directed toward the current task. These researchers admit that they have not modeled this type attention in their interpretation of the ACT-R model. Lovett et al. stated that varying the amount of source activation or determining how it is shared among different components of the current goal could allow for inclusion of this type of attention in the model. However, they did not provide an indication of how this could be completed, only stating that it would be interesting to explore the issue of how people control the allocation of attentional resources instead of the automatic allocation of resources. An unanswered question is, “How does this model account for the decision making processes involved in the control of attention?”

One final aspect of WM that worth requires attention the domain specificity of WM. Daneman and Tardiff [35] argued that individual differences in WM are domain specific and not based on a common or general system. The evidence that they cited for this conclusion is that measures of verbal WM, not visual WM are significant predictors of reading performance. Daneman and Carpenter [7] contended that executive functions do not differ among individuals; however, the storage capacity of WM does differ from person to person. As was discussed earlier, Ericsson and Kintsch [20] describe LTWM as a domain specific component of the memory system where procedures and schemas for complex cognitive tasks are readily accessed from LTM.

Turner and Engle [36] argued that WM is an individual characteristic independent of the nature of the task for which it is applied. As evidence of this conclusion, researchers have found that visual-spatial correlated as highly with reading scores, as did verbal tasks [37]. These findings suggest that WM capacity is not constrained by the specifics of the task. Swanson and Howell [38] found that age-related differences on WM tasks were consistent across verbal and visual-spatial WM tasks. The conclusion then is that visual-spatial and verbal WM tasks access one underlying system.

Before concluding, it is necessary to acknowledge the fact that there are a number of WM models that not specifically mentioned in this paper. However, the models described are of great significance in the evolution of the models of WM. What was revealed in this investigation is that there is a great deal of agreement about what the contents and processes of WM may be. Even those who argue for alternative views of WM have a tendency to use evidence for the opposite views in their arguments. The empirical evidence for WM as a multiple component model is rather convincing.

10. Conclusion

Working memory is the driving force behind complex cognition. Science is just beginning to understand the capacity and limitations that constrain our ability to process information. As this occurs, new questions arise: Is working memory fixed or can we improve our span? Is working memory capacity the key to intelligence? The answer to these types of questions will impact education, work-life and how we deal with old age.

References

[1] J.J. Locke, *An Essay Concerning Human Understanding*, (1975), Peter H. Nidditch (Ed.), Oxford: Clarendon.

[2] W. James, *The Principles of Psychology*, (1890), Oxford, UK: Holt.

[3] N.C. Waugh and D.A. Norman, Primary memory, *Psychological Review*, 72(1965), 89-104.

[4] R.C. Atkinson and R.M. Shiffrin, Human memory: A proposed system and its control processes, In K.W. Spence and J.T. Spence, *The Psychology of Learning and Motivation (Volume 2)*, (1968), 89-195, New York: Academic Press.

[5] F.I.M. Craik and R.S. Lockhart, Levels of processing: A framework for memory research, *Journal of Verbal Learning and Verbal Behavior*, 11(1972), 671-84.

[6] R.H. Logie, *Visuo-Spatial Working Memory*, (1995), Hove, U.K.: Erlbaum.

[7] M. Daneman and P.A. Carpenter, Individual differences in working memory and reading, *Journal of Verbal Learning and Verbal Behavior*, 19(1980), 450-466.

[8] M.A. Just and P.A. Carpenter, A capacity theory of comprehension: Individual differences in working memory, *Psychological Review*, 99(1992), 122-149.

[9] J. King and M.A. Just, Individual differences in syntactic processing: The role of working memory, *Journal of Memory and Language*, 30(1991), 580-602.

[10] A. Baddeley, R. Logie, I. Nimmo-Smith and N. Brereton, Components of fluent reading, *Journal of Memory and Language*, 24(1985), 119-131.

[11] R.D. Case, D.M. Kurland and J. Goldberg, Operational efficiency and the growth of short-term memory span, *Journal of Experimental Child Psychology*, 33(1982), 386-404.

[12] A.R.A. Conway, M.J. Kane, M.F. Bunting, D.Z. Hambrick, O. Wilhelm and R.W. Engle, Working memory span tasks: A methodological review and user's guide, *Psychonomic Bulletin & Review*, 12(5) (2005), 769-786.

[13] M.J. Kane, D.Z. Hambrick, S.W. Tuholski, O. Wilhelm, T.W. Payne and R.W. Engle, The generality of working memory capacity: A latent-variable approach to verbal and visuospatial memory span and reasoning, *Journal of Experimental Psychology: General*, 133(2004), 189-217.

[14] A.D. Baddeley, *Working Memory*, (1986), Oxford: Oxford University Press.

[15] A.D. Baddeley and G.J. Hitch, Working memory, In G.A. Bower (Ed.), *The Psychology of Learning and Motivation: Advances in Research and Theory (Vol. 8)*, (1974), 47-89, New York: Academic Press.

[16] A.D. Baddeley and R.H. Logie, The multiple component model, In A. Miyake and P. Shah (Eds.), *Models of Working Memory: Mechanisms of Active Maintenance and Executive Control*, (1999), 28-61, Cambridge, England: Cambridge University Press.

[17] A.D. Baddeley, Short-term memory for word sequences as a function of acoustic, semantic and formal similarity, *Quarterly Journal of Experimental Psychology*, 18(1996), 362-365.

[18] E.E. Smith and J. Jonides, Working memory: A view from neuroimaging, *Cognitive Psychology*, 33(1997), 5-42.

[19] A.D. Baddeley, The episodic buffer: A new component of working memory? *Trends in Cognitive Sciences*, 4(11) (2000), 417-423.

[20] K.A. Ericsson and W. Kintsch, Long term working memory, *Psychological Review*, 102(1995), 211-245.

[21] S.C. Duff and R.H. Logie, Processing and storage in working memory span, *The Quarterly Journal of Experimental Psychology*, 54a(1) (2001), 31-48.

[22] M. Posner, Attention in cognitive neuroscience: An overview, In M. Gazzaniga (Ed.), *The Cognitive Neurosciences*, (1995), 615-624, Cambridge, MA, MIT Press.

[23] N. Cowan, *Attention and Memory: An Integrated Framework (Oxford Psychology Series, No. 26)*, (1995), New York: Oxford University Press.

[24] N. Cowan, An embedded-processes model of working memory, In A. Miyake and P. Shah (eds.), *Models of Working Memory: Mechanisms of Active Maintenance and Executive Control*, (1999), 62-101, Cambridge, U.K.: Cambridge University Press.

[25] R.A. Conway and R.W. Engle, Working memory and retrieval: A resource-dependent inhibition model, *Journal of Experimental Psychology*, 123(4) (1994), 354-373.

[26] J.R. Anderson, *The Architecture of Cognition*, (1983), Cambridge, Massachusetts, Harvard University Press.

[27] J. Cantor and R.W. Engle, Working memory capacity as long-term memory activation: An individual-differences approach, *Journal of Experimental Psychology: Learning Memory and Cognition*, 19(1993), 1101-1114.

[28] D.D. Wickens, M.J. Moody and R. Dow, The nature and timing of the retrieval process and of interference effects, *Journal of Experimental Psychology*, 110(1981), 1-20.

[29] R.A. Bjork, Retrieval inhibition as an adaptive mechanism in human memory, In H.L.C. Roediger III, *Varieties of Memory and Consciousness*, (1989), 309-330, Hillsdale, NJ, Erlbaum.

[30] R.W. Engle, M.J. Kane and S.W. Tuholski, Individual differences in working memory capacity and what they tell us about controlled attention, general fluid intelligence and functions of the prefrontal cortex, In A. Miyake and P. Shah (Eds.), *Models of Working Memory: Mechanisms of Active Maintenance and Executive Control*, (1999), 102-134, Cambridge, England: Cambridge University Press.

[31] M.C. Lovett, L.M. Reder and C. Lebiere, Modeling working memory in a unified architecture, In A. Miyake and P. Shah (Eds.), *Models of Working Memory: Mechanisms of Active Maintenance and Executive Control*, (1999), 135-182, Cambridge, England: Cambridge University Press.

[32] M.C. Lovett and J.R. Anderson, History of success and current context in problem solving: Combined influences in operator selection, *Cognitive Psychology*, 3(1996), 1168-217.

[33] J.R. Anderson, L.M. Reder and C. Lebiere, Working memory: Activation limitations on retrieval, *Cognitive Psychology*, 30(1996), 221-256.

[34] J.R. Anderson and M.P. Matessa, A production system theory of serial memory, *Psychological Review*, 104(1997), 728-748.

[35] M. Daneman and T. Tardiff, Working memory and reading skill re-examined, In M. Coltheart (Ed.), *Attention and Performance 12: The Psychology of Reading*, (1987), 491-508, Hilldsdale, NJ, Lawrence Earlbaum.

[36] M.L. Turner and R.W. Engle, Is working memory task dependent? *Journal of Memory and Language*, 28(1989), 127-154.

[37] J. Cantor, R.W. Engle and G. Hamilton, Short-term memory, working memory and verbal abilities: How do they relate? *Intelligence*, 15(1991), 229-246.

[38] H.L. Swanson and M. Howell, Working memory, short-term memory and speech rate as predictors of children's reading performance at different ages, *Journal of Educational Psychology*, 93(2001), 720-734.